MULTIPLE RECURRENCE AND CONVERGENCE ALONG THE PRIMES

TREVOR D. WOOLEY AND TAMAR D. ZIEGLER

ABSTRACT. Let $E \subset \mathbb{Z}$ be a set of positive upper density. Suppose that $P_1, P_2, \ldots, P_k \in \mathbb{Z}[X]$ are polynomials having zero constant terms. We show that the set $E \cap (E - P_1(p-1)) \cap \ldots \cap (E - P_k(p-1))$ is non-empty for some prime number p. Furthermore, we prove convergence in L^2 of polynomial multiple averages along the primes.

1. INTRODUCTION

Given a subset E of the integers having positive upper density, the set $E - E$ of differences between pairs of elements of E contains an element of the shape $p - 1$, with p a prime number. This conjecture of Erdős was proved by means of the Hardy-Littlewood (circle) method by Sárközy [18] in a quantitative form which shows that, if $E - E$ contains no shifted prime $p - 1$, then necessarily

$$x^{-1}\mathrm{card}(E \cap [1, x]) \ll \frac{(\log\log\log x)^3(\log\log\log\log x)}{(\log\log x)^2}. \tag{1.1}$$

Subsequent improvements, first by Lucier [15], and most recently by Ruzsa and Sanders [17], show that the function on the right hand side in the conclusion (1.1) may be replaced by $\exp(-c(\log x)^{1/4})$, for some positive absolute constant c. Problems in which one asks for specified constellations of differences between successive terms from a sequence of elements in E, each difference depending on the same shifted prime, have been addressed only very recently. Thus, for example, the problem of exhibiting non-trivial three term arithmetic progressions from E, with common difference a shifted prime, was successfully analysed by Frantzikinakis, Host and Kra [5], with the analogous problem for longer arithmetic progressions conditional on the Inverse Conjecture for Gowers Norms formulated by Green and Tao [8]. Our goal in this paper is the unconditional resolution of a generalisation of these earlier results, an analogue of the Bergelson-Leibman theorem [2], which exhibits a constellation of differences defined by given polynomials whenever these polynomials have zero constant terms.

In order to describe our conclusions, we must introduce some notation, and this we use throughout. We denote by $[N]$ the discrete interval $\{1, \ldots, N\}$ of

natural numbers. Also, we write $|X|$ for the cardinality of a finite set X, and when X is non-empty, we write

$$\mathop{\mathbb{E}}_{n \in X} F(n) = \frac{1}{|X|} \sum_{n \in X} F(n).$$

Given a set of integers E having positive upper density, and polynomials $P_1, \ldots, P_k \in \mathbb{Z}[x]$, we define the return set $R_{P_1, \ldots, P_k}$ by

$$R_{P_1, \ldots, P_k} = \{n \in \mathbb{Z} : E \cap (E - P_1(n)) \cap \ldots \cap (E - P_k(n)) \neq \emptyset\}.$$

Finally, we write $\mathbb{P}$ for the set of prime numbers. It is natural to conjecture that return sets defined by polynomials with zero constant terms contain shifted primes (see, for example, Conjecture 1.1 of [14]). Our first result confirms this conjecture in full generality for the sets $\mathbb{P} \pm 1$ of shifted primes.

Theorem 1.1. *Let E be a set of integers having positive upper density, and let $P_1, \ldots, P_k \in \mathbb{Z}[x]$ satisfy the condition that $P_i(0) = 0$ $(1 \leqslant i \leqslant k)$. Then $R_{P_1, \ldots, P_k} \cap (\mathbb{P} + 1) \neq \emptyset$ and $R_{P_1, \ldots, P_k} \cap (\mathbb{P} - 1) \neq \emptyset$.*

We are also able to establish that polynomial averages converge when restricted to the prime numbers.

Theorem 1.2. *Suppose that $X = (X_0, \mathcal{B}, \mu, T)$ is an invertible measure preserving system. Let $f_1, \ldots, f_k \in L^\infty(X)$, and let $P_1, \ldots, P_k \in \mathbb{Z}[x]$. Then as $N \to \infty$, the averages*

$$\mathop{\mathbb{E}}_{p \in \mathbb{P} \cap [N]} \prod_{j=1}^{k} f_j(T^{P_j(p)} x)$$

converge in $L^2(X)$.

The simplest case of Theorem 1.1 is that in which $k = 1$ and $P_1(n) = n$. As we have already noted in our opening paragraph, this is the case that was successfully considered by Sárközy [18] via the circle method. The convergence of the averages asserted by Theorem 1.2 in this case was apparently first demonstrated by Weirdl [21], and pointwise convergence has also been established (see [4], [20]). In the special case $k = 2$ and $(P_1(n), P_2(n)) = (n, 2n)$, the conclusions of Theorems 1.1 and 1.2 have been proved unconditionally by Frantzikinakis, Host and Kra [5], and subject to the truth of the Inverse Conjecture for Gowers Norms described in [8], this work extends also to any positive integer k and linear polynomials $P_i(n) = in$ $(1 \leqslant i \leqslant k)$[1]. We note, however, that the full conclusions of Theorems 1.1 and 1.2 do not follow from the approach in [5], even if one is prepared to assume the latter Inverse Conjecture. We remark also that Li and Pan [14] have very recently established the case $k = 1$ of Theorem 1.1 when the set of shifted primes is $\mathbb{P} - 1$ (see Corollary 1.1 of [14]).

Rather than attempt to wield control of the prime variable conjecturally made available through Gowers norms, we instead seek control of convergence

---

[1]The Inverse Conjecture for the Gowers norm in the case $k = 4$ is by now known [10], and thus the proof in [5] extends unconditionally to 4-term arithmetic progressions.

through a variable from the set E, switching the roles of this variable and the prime. Such a strategy, in which for less well controlled aspects of an analysis one may crudely count prime variables by inclusion in larger well-behaved subsets of the integers, is reasonably familiar to practitioners of the circle method and sieve theory. The mechanism which makes this switching of roles effective is the use of the local Gowers norms introduced in [19]. This allows us to assume that the set E possesses extra structure, namely a *nilstructure*. With this information in hand, we are able to apply the recent work of Green and Tao [9], showing that the Möbius function is orthogonal to (polynomial) nilsequences, in combination with the Leibman structure theorem for multivariable polynomial averages [11] in order to deliver the conclusions of Theorems 1.1 and 1.2.

It seems likely that our methods could be adapted to handle modifications of Theorems 1.1 and 1.2 in which the polynomials in $\mathbb{Z}[x]$ are replaced by general integer-valued polynomials. Indeed, even the restriction to polynomials having vanishing constant terms might be weakened through a modification of the hypotheses of Theorem 1.1 to accommodate jointly intersective polynomials (see [3] for the relevant ideas).

We have recorded a number of notational and technical preliminaries relating to the ergodic theory that we employ in two appendices at the end of this paper. Readers not already aficionados of the subject area would be well-advised to peruse this material before continuing further. In particular, we take this opportunity to emphasise that throughout this paper, whenever we refer to a measure preserving system, we implicitly assume this system to be *invertible*. In section 2 we outline our approach to the central problem of the paper. We consider the prime return set $R_{P_1,\ldots,P_k}$ in section 3, providing the details of the proof of Theorem 1.1. Section 4 is devoted to the convergence of polynomial averages restricted to the primes, leading to the proof of Theorem 1.2.

The authors are grateful to the referees of this paper for their careful reading, detailed comments, and the consequent improvement in our exposition.

2. Outline of proof

We begin by considering a k-tuple of polynomials $\mathbf{P} = (P_1, \ldots, P_k)$, and a set E having upper density exceeding some positive number δ. Our first step is to translate the question on the prime return set into an ergodic theoretic one via Furstenberg's correspondence principle. Thus we replace the set E by a measurable set A, of measure $\mu(A) > \delta$, in a probability measure preserving system $\mathrm{X} = (X, \mathcal{B}, \mu, T)$. By a uniform version of the Bergelson-Leibman theorem (see Theorem 3.9 below), there is a positive number $c(\delta)$ with the property that for any natural number W, one has

$$\lim_{N \to \infty} \mathbb{E}_{n \leqslant N} \mu(T^{-P_1(Wn)}A \cap \ldots \cap T^{-P_k(Wn)}A) > c(\delta).$$

Here, we emphasise that the number $c(\delta)$ depends on δ, as well as the polynomials $P_1, \ldots, P_k$, but is independent of A and W.

The ordered polynomial system $\mathcal{P} = \{P_1, \ldots, P_k\}$ determines, via PET induction, the number of steps $l(\mathcal{P})$ that one must take, by repeated application of the Cauchy-Schwarz inequality, to obtain a parallelepiped system of polynomials independent of the parameter n. This in turn determines the sieve level $R = N^\eta$, also independent of W, via the condition $\eta < 2^{-3-l(\mathcal{P})}$. All estimates henceforth depend implicitly on δ and η.

In our next step, we take w to be a slowly growing function of N, and put

$$W = \prod_{\substack{p < w \\ p \in \mathbb{P}}} p.$$

We will sometimes need to fix w (very large) and take N much larger, and for this reason it is useful to adopt the following convention concerning Landau's o-notation within this paper. As usual, when a quantity approaches zero as the main parameter N approaches infinity, we shall say that this quantity is $o(1)$. We denote by $o_{w \to \infty}(1)$ any quantity that approaches zero as $w \to \infty$. Finally, we denote by $o_w(1)$ any quantity that, with w fixed, approaches zero as $N \to \infty$.

Next, let b be an integer with $(b, W) = 1$. Perhaps it is worth noting that, when it comes to establishing Theorem 1.2 in section 4, we must consider all possible values of b. However, for the proof of Theorem 1.1 in section 3, it transpires that the only values of b of interest are ± 1 (see the discussion surrounding (3.4) below). We define the function $\widetilde{\Lambda}_{w,b}(n)$ by putting

$$\widetilde{\Lambda}_{w,b}(n) = \frac{\phi(W)}{W} \log R = \eta \frac{\phi(W)}{W} \log N,$$

when $Wn + b$ is a prime number[2], and otherwise by putting $\widetilde{\Lambda}_{w,b}(n) = 0$. Here, as usual, we write $\phi(W)$ for the Euler totient, so that $\phi(W) = \prod_{p<w}(p-1)$. In [19], an enveloping sieve argument is applied to show that there exists a function $\nu_{w,b}(n)$ with the property that $\widetilde{\Lambda}_{w,b}(n) \leqslant \nu_{w,b}(n)$, so that $\widetilde{\Lambda}_{w,b}$ is pointwise bounded by $\nu_{w,b}$, and

$$\|\nu_{w,b} - 1\|_{V_{\mathcal{P}}} = o_{w \to \infty}(1).$$

Although we defer until later the definition of the norm here, it may be helpful to note that it is similar to a Gowers norm, though with shift sizes short with respect to N, but larger than the sieve level R. We remark that our use of notation differs from that in [19], owing to the simpler nature of the polynomials in question, as well as the absence of scaling issues which obviates the need for the full structure theorem proved in [19].

We examine the average

$$\mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \mu(T^{-P_1(Wn)} A \cap \ldots \cap T^{-P_k(Wn)} A), \qquad (2.1)$$

---

[2]In the detailed account of our argument in section 3, we make the additional technical restriction that $\widetilde{\Lambda}_{w,b}(n)$ is thus defined only when $n \in [\frac{1}{2}N]$. The straightforward complications associated with this constraint are best ignored in the present outline.

and make use of the majorant $\nu_{w,b}$ of $\widetilde{\Lambda}_{w,b}$ to compare it to the related average

$$\mathbb{E}_{n \leqslant N} \eta\mu(T^{-P_1(Wn)}A \cap \ldots \cap T^{-P_k(Wn)}A), \tag{2.2}$$

which we already know to exceed $\eta c(\delta)$. Our aim is to show that the difference between these averages is $o_{w\to\infty}(1)$. This we achieve in two steps. The parameter $l(\mathcal{P})$ determines a factor $Z_{l(\mathcal{P})}(\mathrm{X})$ having the structure of an $(l(\mathcal{P})-1)$-step nilsystem, this system being independent of W. In the first step, we show that when f_i is orthogonal to $Z_{l(\mathcal{P})}(\mathrm{X})$ for some index i with $1 \leqslant i \leqslant k$, or equivalently, when $\pi : \mathrm{X} \to Z_{l(\mathcal{P})}(\mathrm{X})$ is the factor map and $\pi_* f_i = 0$, then

$$\mathbb{E}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \int \prod_{j=1}^{k} T^{P_j(Wn)} f_j(x)\, d\mu = o_w(1) + o_{w\to\infty}(1).$$

As usual, here and throughout, we write $Tf(x)$ for $f(Tx)$. We then decompose the characteristic function on A by means of the trivial relation $1_A = \pi^*\pi_* 1_A + (1_A - \pi^*\pi_* 1_A)$. Then $\pi_*(1_A - \pi^*\pi_* 1_A) = 0$, and thus

$$\mathbb{E}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n)\mu(T^{-P_1(Wn)}A \cap \ldots \cap T^{-P_k(Wn)}A)$$

$$= \mathbb{E}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \int \prod_{j=1}^{k} (T^{P_j(Wn)}\pi_* 1_A(x))\, d\pi_*\mu + o_w(1) + o_{w\to\infty}(1).$$

This allows us to reduce to the situation in which the system X is an $l(\mathcal{P})$-step pro-nilsystem. In fact, technically speaking, we replace A by the non-negative function $\pi_* 1_A$, which has integral against $\pi_*\mu$ exceeding δ. We note that the universality of the constant $c(\delta)$ applies for any such function.

We make an additional reduction to the case in which f is defined on a nilsystem $(G/\Gamma, \mathcal{B}, \mu, T)$. This is achieved by means of an approximation in L^2, and is independent of w. If this system is disconnected, then it can be decomposed into a union of some finite number, J, of components $\{\mathrm{X}_i\}_{i=1}^{J}$ having the property that $T^J : \mathrm{X}_i \to \mathrm{X}_i$ is totally ergodic for $1 \leqslant i \leqslant J$.

We now follow the argument of [8]. We replace $\widetilde{\Lambda}_{w,b}(n)$ by the function

$$\Lambda_{w,b}(n) = \eta \frac{\phi(W)}{W} \Lambda(Wn + b),$$

in which Λ denotes the classical von Mangoldt function. We then decompose Λ by means of a Möbius identity into the shape $\Lambda^{\sharp} + \Lambda^{\flat}$, corresponding to an associated smooth decomposition of the identity function $\chi(x) = x$ in the shape $\chi = \chi^{\sharp} + \chi^{\flat}$, with $\Lambda^{\sharp}$ associated to small divisors and $\Lambda^{\flat}$ associated to large divisors, just as in [8]. Observe next that for any Lipschitz function f, the expression

$$\prod_{j=1}^{k} (T^{P_j(Wn)} f(x))$$

is a polynomial nilsequence on $(G/\Gamma)^k$. As in [8], we show that the contribution arising from the term

$$\mathop{\mathbb{E}}_{n \leqslant N} \Big(\frac{\phi(W)}{W} \Lambda^\sharp(Wn + b) - 1 \Big) \prod_{j=1}^{k} (T^{P_j(Wn)} f(x))$$

is negligible. The estimate of the contribution arising from the term corresponding to $\Lambda^\flat$ follows from Theorem 1.1 of [9], which asserts that the Möbius function is orthogonal to polynomial nilsequences with bounds that depend only on the degree of the polynomial and not on the polynomial itself.

Our goal of showing that the averages (2.1) and (2.2) are asymptotically equal is completed by combining the results of the last paragraph, and this completes our outline of the proof.

3. Prime return sets

Our objective in this section is the proof of Theorem 1.1. We begin with a discussion of the pseudorandom measures employed in the sketch of the argument provided in the previous section.

3.1. **Pseudorandom measures.** We first define a normalised counting function for prime numbers, with a smoothing weight designed to flatten distribution across a subset of residue classes. Let η be a positive number with $\eta < 2^{-3-l(\mathcal{P})}$, and put $R = N^\eta$. Define the function $\widetilde{1} : [N] \to \mathbb{R}^+$ by putting $\widetilde{1}(x) = 1$ when $x \in [\frac{1}{2}N]$, and otherwise by taking $\widetilde{1}(x) = 0$. In addition, define $\widetilde{\Lambda}_{w,b} : [N] \to \mathbb{R}^+$ by setting

$$\widetilde{\Lambda}_{w,b}(x) = \frac{\phi(W)}{W} \log R, \tag{3.1}$$

when $x \in [\frac{1}{2}N]$ and $Wx + b \in \mathbb{P}$, and otherwise by taking $\widetilde{\Lambda}_{w,b}(x) = 0$. Here, we choose to identify $[\frac{1}{2}N]$ with a subset of $\mathbb{Z}/N\mathbb{Z}$, in the usual manner. We remark that the function $\widetilde{\Lambda}_{w,b}(x)$ is a modification of the classical von Mangoldt function $\Lambda(x)$. The use of $\log R$ in place of $\log N$, as a normalising factor, is necessary in order to bound $\widetilde{\Lambda}$ pointwise by the pseudorandom measure ν shortly to be defined. The ratio η between $\log R$ and $\log N$ reflects the relative density between the primes, and the almost primes occurring implicitly within our argument.

An application of the Prime Number Theorem in arithmetic progressions with error term (see, for example, Corollary 11.21 of [16]) reveals that when b and W are coprime, one has

$$|\{x \in [\tfrac{1}{2}N] : Wx + b \in \mathbb{P}\}| \gg \frac{W}{\phi(W)} \frac{N}{\log N}.$$

It follows that $\widetilde{\Lambda}_{w,b}$ has relatively large mean, namely

$$\mathop{\mathbb{E}}_{n \in [N]} \widetilde{\Lambda}_{w,b} \gg \eta.$$

Before announcing the key properties of the pseudorandom measure employed in our argument, we must record some definitions. The first definition, of a measure, comes from Definition 6.1 of [8].

Definition 3.1. A *measure* is a non-negative function $\nu_w : [N] \to \mathbb{R}^+$ with the total mass estimate

$$\underset{n\in[N]}{\mathbb{E}} \, \nu_w = 1 + o_{w\to\infty}(1), \tag{3.2}$$

and such that for each positive number ε, one has the crude pointwise bound $\nu_w = O_\varepsilon(N^\varepsilon)$.

Next we define polynomial norms analogous to Gowers norms.

Definition 3.2. Let a be a function from $\mathbb{Z}$ into $\mathbb{C}$ supported in $[N]$. When k is a non-negative integer, we define the V_k-*norm* of a to be the quantity $\|a\|_{V_k}$ defined via the relation

$$\|a\|_{V_k}^{2^k} = \underset{n\leqslant N}{\mathbb{E}} \underset{\substack{m_1,\ldots,m_k\leqslant\sqrt{N}\\ m_1',\ldots,m_k'\leqslant\sqrt{N}}}{\mathbb{E}} \prod_{\boldsymbol{\omega}\in\{0,1\}^k} a^{\boldsymbol{\omega}}(n + \boldsymbol{\omega}\cdot\mathbf{m} + (\mathbf{1}-\boldsymbol{\omega})\cdot\mathbf{m}').$$

Here, we write $\mathbf{1}$ for the vector $(1,1,\ldots,1)$, and we put $a^{\boldsymbol{\omega}} = a$ when $\sum_{i=1}^k \omega_i \equiv 0 \pmod 2$, and otherwise we put $a^{\boldsymbol{\omega}} = \overline{a}$. Also, when $\mathcal{P} = \{P_1,\ldots,P_k\}$ is a standard polynomial system with parallelepiped order $l(\mathcal{P})$, we define the $V_{\mathcal{P}}$-norm of the function a by $\|a\|_{V_{\mathcal{P}}} = \|a\|_{V_{l(\mathcal{P})+1}}$.

Observe that

$$\|a\|_{V_1}^2 = \underset{n\leqslant N}{\mathbb{E}} \left| \underset{m\leqslant\sqrt{N}}{\mathbb{E}} a(n+m) \right|^2 \geqslant 0,$$

so that the definition of the V_k-norm makes sense when $k = 1$. For larger values of k, such follows from the following lemma, which records two simple properties of the V_k-norm useful in our subsequent deliberations.

Lemma 3.3. *Let a be a function from $\mathbb{Z}$ into $\mathbb{C}$ supported in $[N]$. When k is a non-negative integer and $0 < \gamma \leqslant 2^k$, one has*

$$\underset{m,m'\leqslant\sqrt{N}}{\mathbb{E}} \|a(n+m)\overline{a}(n+m')\|_{V_k}^{\gamma} \leqslant \|a\|_{V_{k+1}}^{2\gamma},$$

with equality when $\gamma = 2^k$. If, moreover, the function a has the property that for each positive number ε, one has the pointwise bound $a(n) = O_\varepsilon(N^\varepsilon)$, then

$$\left| \underset{n\leqslant N}{\mathbb{E}} \, a(n) \right| \leqslant \|a\|_{V_1} + o(1).$$

Proof. The first claim follows at once from the definition of the V_k-norm, since by Hölder's inequality one has

$$\underset{m,m'\leqslant\sqrt{N}}{\mathbb{E}} \|a(n+m)\overline{a}(n+m')\|_{V_k}^{\gamma} \leqslant \left(\underset{m,m'\leqslant\sqrt{N}}{\mathbb{E}} \|a(n+m)\overline{a}(n+m')\|_{V_k}^{2^k} \right)^{\gamma 2^{-k}},$$

and the expectation within parentheses on the right hand side here is equal to

$$\mathbb{E}_{\substack{n \leqslant N \\ m_0,\ldots,m_k \leqslant \sqrt{N} \\ m_0',\ldots,m_k' \leqslant \sqrt{N}}} \prod_{\boldsymbol{\omega} \in \{0,1\}^{k+1}} a^{\boldsymbol{\omega}}(n + \boldsymbol{\omega} \cdot \mathbf{m} + (\mathbf{1} - \boldsymbol{\omega}) \cdot \mathbf{m}') = \|a\|_{V_{k+1}}^{2^{k+1}}.$$

The final conclusion of the lemma is essentially a consequence of the van der Corput lemma, as in the proof of Lemma A.1 of [19], though here we are more precise and do not restrict to real functions. Observe that, as a consequence of our hypotheses concerning $a(n)$, one has

$$\mathbb{E}_{n \leqslant N} a(n) = \mathbb{E}_{m \leqslant \sqrt{N}} \mathbb{E}_{n \leqslant N} a(n + m) + O(N^{\varepsilon - 1/2}).$$

Interchanging the order of summation, an application of Cauchy's inequality yields

$$\left| \mathbb{E}_{n \leqslant N} a(n) \right|^2 \leqslant \mathbb{E}_{n \leqslant N} \mathbb{E}_{m,m' \leqslant \sqrt{N}} \prod_{\omega \in \{0,1\}} a^{\omega}(n + \omega m + (1 - \omega)m') + O(N^{2\varepsilon - 1/2})$$

$$= \|a\|_{V_1}^2 + o(1).$$

The desired conclusion is now immediate. $\qquad\qquad\square$

The following theorem is essentially equivalent to Theorem 3.18 of [19], and demonstrates the existence of a pseudorandom majorant[3].

Theorem 3.4. *Let $\mathcal{P}$ be a standard polynomial system, and let $\eta = 2^{-3-l(\mathcal{P})}$. Then there exists a measure $\nu_{w,b}$ with the property that the function $\widetilde{\Lambda}_{w,b}$ defined in (3.1) enjoys the pointwise bound $0 \leqslant \widetilde{\Lambda}_{w,b} \leqslant \nu_{w,b}$, and further*

$$\|\nu_{w,b} - 1\|_{V_{\mathcal{P}}} = o_{w \to \infty}(1). \tag{3.3}$$

We note that in [19], the parameter w is concretely fixed to be of order $\log\log\log N$. In present circumstances, meanwhile, we prefer to think of w as very (very) large, but constant, since in the ergodic convergence results we do not have uniformity in w. To clarify the dependence on w, we use both the notations $o_w(1)$ and $o_{w \to \infty}(1)$, as defined in section 2.

3.2. **Translation to the ergodic world.** We open the main thrust of our argument by translating the basic question to an ergodic theoretic setting. We achieve this goal by means of the Furstenberg Correspondence Principle (see, for example, Furstenberg [6]).

Lemma 3.5. *Let E be a set of positive upper density in $\mathbb{Z}$. Then there exists a measure preserving system $\mathrm{X} = (X, \mathcal{B}, \mu, T)$, and an element A of $\mathcal{B}$ with $\mu(A) > 0$, with the property that when*

$$\mu(A \cap T^{-n_1}A \cap \ldots \cap T^{-n_k}A) > 0,$$

then

$$E \cap (E - n_1) \cap \ldots \cap (E - n_k) \neq \emptyset.$$

---

[3] In modern language, the measure whose existence is asserted by Theorem 3.4 is described as a *pseudorandom measure*, by virtue of the property (3.3)

Making use of ergodic decomposition, it follows as a corollary of this conclusion that in order to prove Theorem 1.1, it suffices to establish the following ergodic theoretic version of this theorem.

Theorem 3.6. *Suppose that* $X = (X, \mathcal{B}, \mu, T)$ *is an ergodic measure preserving system, and let* $P_1, \ldots, P_k \in \mathbb{Z}[x]$ *satisfy the condition* $P_i(0) = 0$ $(1 \leqslant i \leqslant k)$. *In addition, suppose that* $A \in \mathcal{B}$ *satisfies the condition* $\mu(A) > 0$. *Let*

$$S_{P_1,\ldots,P_k} = \{n \in \mathbb{Z} : \mu(A \cap T^{-P_1(n)}A \cap \ldots \cap T^{-P_k(n)}A) > 0\}.$$

Then

$$S_{P_1,\ldots,P_k} \cap (\mathbb{P} + 1) \neq \emptyset \quad and \quad S_{P_1,\ldots,P_k} \cap (\mathbb{P} - 1) \neq \emptyset.$$

As in many other recurrence results, it is easier to show that the set $S_{P_1,\ldots,P_k} \cap (\mathbb{P} \pm 1)$ is large than merely showing that it is not empty. In particular, it suffices to show that for any integer b with $(b, W) = 1$, one has

$$\mathop{\mathbb{E}}_{\substack{n \leqslant N \\ Wn+b \in \mathbb{P}}} \mu(A \cap T^{-P_1(Wn)}A \cap \ldots \cap T^{-P_k(Wn)}A) > 0. \tag{3.4}$$

Notice here that we have no useful control over W. However, since $(\pm 1, W) = 1$, it follows from the Siegel-Walfisz theorem (see, for example, Corollary 11.21 of [16]) that for large enough values of N and $b = \pm 1$, the expectation in (3.4) is taken over a non-empty set. Hence, the lower bound (3.4) is sufficient to establish Theorem 3.6. On the other hand, the set $\mathbb{P} - 2$ is *not* a return set for polynomial averages.

The next lemma is classical.

Lemma 3.7. *Suppose that* $|a_n| < 1$ *for each integer* n. *Then one has*

$$\left| \mathop{\mathbb{E}}_{\substack{n \leqslant N \\ Wn+b \in \mathbb{P}}} a_{Wn+b} - \mathop{\mathbb{E}}_{n \leqslant N} \frac{\phi(W)}{W}\Lambda(Wn + b)a_{Wn+b} \right| = o_w(1).$$

As a consequence of this result, one may replace the average on the left hand side of (3.4) by a weighted average, wherein the weights are given by a modified von Mangoldt function. This conclusion we summarise in the next lemma.

Lemma 3.8. *Suppose that* $\mu(A) > \delta$, *for some positive number* δ. *Then, in order to establish the lower bound (3.4), it suffices to show that*

$$\mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n)\mu(A \cap T^{-P_1(Wn)}A \cap \ldots \cap T^{-P_k(Wn)}A) \gg_\delta 1 + o_w(1) + o_{w\to\infty}(1).$$

Equivalently, writing $1_A(x)$ *for the characteristic function of the set* A, *it suffices to show that*

$$\mathop{\mathbb{E}}_{n \leqslant N} \int \widetilde{\Lambda}_{w,b}(n)1_A(x) \prod_{j=1}^{k}(T^{P_j(Wn)}1_A(x)) \, d\mu \gg_\delta 1 + o_w(1) + o_{w\to\infty}(1). \tag{3.5}$$

In order confirm (3.5), we require two additional results. The first treats an analogous situation in which the von Mangoldt weights are absent, a quantitative version of the Polynomial Szemerédi theorem.

Theorem 3.9. *With the notation and assumptions of the previous section, suppose that $\delta > 0$, and let $g : X \to \mathbb{R}$ be any function obeying the pointwise bound $0 \leqslant g \leqslant 1 + o(1)$, together with the mean bound $\int_X g \, d\mu \geqslant \delta - o(1)$. Then we have*

$$\mathop{\mathbb{E}}_{n \leqslant N} \widetilde{1}(n) \int_X g(x) \prod_{j=1}^{k} (T^{P_j(Wn)} g(x)) \, d\mu \geqslant c(\delta) - o_w(1),$$

where $c(\delta)$ is a positive number depending on δ and $P_1, \ldots, P_k$, but independent of W.

Proof. This follows from Theorem 3.2 of [19]. $\square$

We also require the following structure theorem, due to Leibman [11], identifying nilsystems as characteristic factors for multivariate polynomial multiple averages.

Theorem 3.10. *Suppose that $\mathrm{X} = (X, \mathcal{B}, \mu, T)$ is an ergodic measure preserving system. Let $Q_1, \ldots, Q_s \in \mathbb{Z}[x_1, \ldots, x_m]$ be polynomials. In addition, let $\mathcal{Q}$ denote $\{Q_1, \ldots, Q_s\}$. Then there exists a factor $\mathrm{Y} = (Y, \mathcal{D}, \nu, S)$ of X, with $\pi : \mathrm{X} \to \mathrm{Y}$ as the factor map, and an integer $d(\mathcal{Q})$, such that:*

(i) *the system Y has the structure of an inverse limit of $d(\mathcal{Q})$-step nilsystems, and*

(ii) *the average difference*

$$\mathop{\mathbb{E}}_{\mathbf{m} \in \mathcal{M}} \prod_{j=1}^{s} T^{Q_j(W\mathbf{m})} f_j - \pi^* \mathop{\mathbb{E}}_{\mathbf{m} \in \mathcal{M}} \prod_{i=1}^{s} S^{Q_j(W\mathbf{m})} \pi_* f_j$$

is $o_w(1)$ in $L^2(X)$. Here, we have written $\mathcal{M}$ for $[M_1] \times \ldots \times [M_m]$, and the convergence is as $M_1, \ldots, M_m \to \infty$.

Note that the rate of convergence in this theorem may depend on w. What is crucial is that the integer $d(\mathcal{Q})$ is independent of w.

We at last come to the result of this section which does the heavy lifting in our argument. This provides a conclusion on orthogonality to nilsystems.

Proposition 3.11. *Suppose that X is an ergodic measure preserving system. Let $f_1, \ldots, f_k \in L^\infty(X)$ be functions satisfying the condition $\|f_j\|_\infty \leqslant L$ ($1 \leqslant j \leqslant k$). Then there exists a factor Y of X, with $\pi : \mathrm{X} \to \mathrm{Y}$ as the factor map, and an integer $d(\mathcal{P})$, such that*

(i) *the system Y has the structure of an inverse limit of $d(\mathcal{P})$-step nilsystems, and*

(ii) *if, for some index i, one has $\pi_* f_i = 0$, then*

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn)} f_j(x) \, d\mu \right| = o_{L,w}(1) + o_{L,w \to \infty}(1).$$

Proof. The expression which we seek to estimate is

$$\mathcal{T} = \left| \int \underset{n \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn)} f_j(x) \, d\mu \right|^2.$$

Observe first that, by the invariance of the measure μ under the action of T, it follows that for each positive number M, one has

$$\mathcal{T} = \left| \int \underset{n \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \underset{l \leqslant M}{\mathbb{E}} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \, d\mu \right|^2.$$

Consequently, by applying the Cauchy-Schwarz inequality in combination with the triangle inequality, one obtains

$$\mathcal{T} \leqslant \int \left(\underset{n \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \Big| \underset{l \leqslant M}{\mathbb{E}} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \Big| \right)^2 d\mu.$$

By Theorem 3.4, the modified von Mangoldt function $\widetilde{\Lambda}_{w,b}(n)$ is pointwise bounded by the pseudorandom majorant $\nu_{w,b}(n)$, and hence we may replace the former by the latter in the last upper bound for $\mathcal{T}$. Proceeding first in this way, and then applying the Cauchy-Schwarz inequality once again, we deduce that

$$\mathcal{T} \leqslant \int \left(\underset{n \leqslant N}{\mathbb{E}} \, \nu_{w,b}(n) \Big| \underset{l \leqslant M}{\mathbb{E}} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \Big| \right)^2 d\mu$$

$$\leqslant \int \left(\underset{n \leqslant N}{\mathbb{E}} \, \nu_{w,b}(n) \right) \left(\underset{n \leqslant N}{\mathbb{E}} \, \nu_{w,b}(n) \Big| \underset{l \leqslant M}{\mathbb{E}} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \Big|^2 \right) d\mu.$$

Our goal in the remainder of the proof is to establish that the integral

$$\mathcal{U} = \int \underset{n \leqslant N}{\mathbb{E}} \, \nu_{w,b}(n) \Big| \underset{l \leqslant M}{\mathbb{E}} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \Big|^2 d\mu \tag{3.6}$$

satisfies

$$\mathcal{U} = o_{L,w}(1) + o_{L,w \to \infty}(1). \tag{3.7}$$

Since equation (3.2) provides the estimate

$$\underset{n \leqslant N}{\mathbb{E}} \, \nu_{w,b}(N) = 1 + o_{w \to \infty}(1)$$

for the average of the measure $\nu_{w,b}(n)$, it follows from our earlier estimate for $\mathcal{T}$ together with (3.6) and (3.7) that

$$\mathcal{T} \leqslant (1 + o_{w \to \infty}(1))\mathcal{U} = o_{L,w}(1) + o_{L,w \to \infty}(1),$$

and this suffices to complete the proof of the theorem.

We now focus on (3.6), expanding the square in the integrand to obtain

$$\mathcal{U} = \int \mathop{\mathbb{E}}_{n \leqslant N} \nu_{w,b}(n) \mathop{\mathbb{E}}_{l,m \leqslant M} \prod_{j=1}^{k} T^{P_j(Wn)+Wl} f_j(x) \prod_{j=1}^{k} T^{P_j(Wn)+Wm} \overline{f}_j(x) \, d\mu$$

$$= \int \mathop{\mathbb{E}}_{n \leqslant N} \nu_{w,b}(n) \mathop{\mathbb{E}}_{l,m \leqslant M} \prod_{j=1}^{k} T^{P_j(Wn)} (f_j(x) T^{W(m-l)} \overline{f}_j(x)) \, d\mu.$$

Consider the average

$$\int \mathop{\mathbb{E}}_{n \leqslant N} \mathop{\mathbb{E}}_{l,m \leqslant M} \prod_{j=1}^{k} T^{P_j(Wn)} (f_j(x) T^{W(m-l)} \overline{f}_j(x)) \, d\mu.$$

Take M to be a real number with $M = N^{O(1)}$. In addition, write

$$Q_{2i-1}(n, m, l) = P_i(n) \quad \text{and} \quad Q_{2i}(n, m, l) = P_i(n) + m - l \quad (1 \leqslant i \leqslant k),$$

and put $\mathcal{Q} = \{Q_1, \ldots, Q_{2k}\}$. Let Y be the factor supplied by Theorem 3.10 associated with $\mathcal{Q}$. Then if for some i one has $\pi_* f_i = 0$, then from the latter theorem it follows that the above average is $o_{L,w}(1)$.

In view of the above discussion, it suffices to show that for any continuous bounded functions $g_1, \ldots, g_k$ with $\|g_i\|_\infty \leqslant L^2$, one has

$$\int \mathop{\mathbb{E}}_{n \leqslant N} (\nu_{w,b}(n) - 1) \prod_{j=1}^{k} T^{P_j(Wn)} (g_j(x) T^{W(m-l)} \overline{g}_j(x)) \, d\mu = o_{L,w}(1) + o_{L,w \to \infty}(1).$$

We establish the latter by applying PET induction to show that, whenever a is a function from $\mathbb{Z}$ into $\mathbb{C}$ supported in $[N]$, and satisfying $a(n) = O_\varepsilon(N^\varepsilon)$ for every $\varepsilon > 0$, then one has

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} a(n) g_0(x) \prod_{j=1}^{k} T^{P_j(Wn)} g_j(x) \, d\mu \right| \ll_L \|a(n)\|_{V_\mathcal{P}} + o(1). \tag{3.8}$$

The procedure here is very similar to that applied in [19], but unfortunately it does not fit precisely into the framework of the latter. We therefore repeat the process in the present context. The trick is to insert some additional averaging by means of a parameter M of order $\sqrt{N}$. An important observation, in this context, is that since the polynomials may be supposed distinct, with zero constant terms, then the system $\{P_1, \ldots, P_k\}$ may be reordered in such a way that we obtain a standard system.

We first establish the case in which $\mathcal{P}$ is a standard linear system. Thus we suppose that $\mathcal{P} = \{P_1, \ldots, P_k\}$ is a standard linear system, and prove by induction on k that

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} a(n) g_0(x) \prod_{j=1}^{k} T^{P_j(Wn)} g_j(x) \, d\mu \right| \ll_L \|a(n)\|_{V_{k+1}} + o(1).$$

For $k = 1$, we must estimate the absolute value of the integral

$$\mathcal{I}_1^* = \int \mathop{\mathbb{E}}_{n \leqslant N} a(n) g_0(x) T^{P_1(Wn)} g_1(x) \, d\mu.$$

We observe first that $\mathcal{I}_1^* = \mathcal{I}_1 + o_L(1)$, where we have written

$$\mathcal{I}_1 = \mathop{\mathbb{E}}_{n \leqslant N} \int g_0(x) \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) T^{P_1(W(n+m))} g_1(x) \, d\mu.$$

By the Cauchy-Schwarz inequality, one has

$$\mathcal{I}_1 \ll_L \mathop{\mathbb{E}}_{n \leqslant N} \left(\int \left| \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) T^{P_1(W(n+m))} g_1(x) \right|^2 d\mu \right)^{1/2}.$$

By another application of the Cauchy-Schwarz inequality, we obtain the upper bound

$$\mathcal{I}_1^2 \ll_L \mathop{\mathbb{E}}_{n \leqslant N} \int \left| \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) T^{P_1(W(n+m))} g_1(x) \right|^2 d\mu$$

$$= \mathop{\mathbb{E}}_{n \leqslant N} \int \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} a(n+m) \overline{a}(n+m') T^{P_1(W(n+m))} g_1(x) T^{P_1(W(n+m'))} \overline{g}_1(x) \, d\mu.$$

Consequently, by the triangle inequality,

$$\mathcal{I}_1^2 \ll_L \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} \left| \mathop{\mathbb{E}}_{n \leqslant N} a(n+m) \overline{a}(n+m') \right| \left| \int g_1(x) T^{P_1(W(m'-m))} \overline{g}_1(x) \, d\mu \right|.$$

Thus, on applying Lemma 3.3 and making yet another application of the Cauchy-Schwarz inequality, we deduce that

$$\mathcal{I}_1^4 \ll_L \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} (\|a(n+m) \overline{a}(n+m')\|_{V_1} + o(1))^2 \ll \|a\|_{V_2}^4 + o(1).$$

This confirms the inductive hypothesis when $k = 1$.

Suppose now that $K > 1$, and the inductive hypothesis holds for $k < K$. In this case we evaluate the expression

$$\mathcal{I}_K^* = \int \mathop{\mathbb{E}}_{n \leqslant N} a(n) g_0(x) \prod_{j=1}^{K} T^{P_j(Wn)} g_j(x) \, d\mu. \tag{3.9}$$

As before, we first obtain the relation $\mathcal{I}_K^* = \mathcal{I}_K + o_L(1)$, where

$$\mathcal{I}_K = \mathop{\mathbb{E}}_{n \leqslant N} \int g_0(x) \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) \prod_{j=1}^{K} T^{P_j(W(n+m))} g_j(x) \, d\mu.$$

Next, following an application of the Cauchy-Schwarz inequality, we obtain

$$\mathcal{I}_K \ll_L \mathop{\mathbb{E}}_{n \leqslant N} \left(\int \left| \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) \prod_{j=1}^{K} T^{P_j(W(n+m))} g_j(x) \right|^2 d\mu \right)^{1/2}.$$

A further application of the Cauchy-Schwarz inequality leads to the relation

$$\mathcal{I}_K^2 \ll_L \mathop{\mathbb{E}}_{n \leqslant N} \int \Big| \mathop{\mathbb{E}}_{m \leqslant \sqrt{N}} a(n+m) \prod_{j=1}^K T^{P_j(W(n+m))} g_j(x) \Big|^2 d\mu$$

$$= \mathop{\mathbb{E}}_{n \leqslant N} \int \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} a(n+m)\overline{a}(n+m')$$

$$\times \prod_{j=1}^K T^{P_j(W(n+m))} g_j(x) \prod_{j=1}^K T^{P_j(W(n+m'))} \overline{g}_j(x) \, d\mu. \qquad (3.10)$$

Next, owing to the invariance of μ under the action of T, we see that

$$\mathcal{I}_K^2 \ll_L \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} \int \mathop{\mathbb{E}}_{n \leqslant N} a(n+m)\overline{a}(n+m')$$

$$\times g_1^{(m,m')}(x) \prod_{j=2}^K T^{(P_j-P_1)(Wn)} g_j^{(m,m')}(x) \, d\mu,$$

where

$$g_j^{(m,m')}(x) = T^{(P_j-P_1)(Wm)} g_j(x) T^{P_j(Wm')-P_1(Wm)} \overline{g}_j(x) \quad (1 \leqslant j \leqslant K).$$

As a consequence of the inductive hypothesis, we therefore deduce by means of Lemma 3.3 that

$$\mathcal{I}_K^2 \ll_L \mathop{\mathbb{E}}_{m,m' \leqslant \sqrt{N}} (\|a(n+m)\overline{a}(n+m')\|_{V_K} + o(1)) \ll \|a(n)\|_{V_{K+1}}^2 + o(1).$$

This confirms the inductive hypothesis, for standard linear systems, when $k = K$. The inductive hypothesis consequently holds for all standard linear systems.

We now apply the PET induction scheme so as to reduce the general case to one in which the system $\mathcal{P}$ is standard and linear. We proceed by induction on the weight $w(\mathcal{P})$ of the polynomial system $\mathcal{P}$. Suppose that the desired conclusion holds for every standard polynomial system $\mathcal{P}$ with weight $w(\mathcal{P}) < w$. Since we have already established the desired conclusion for every standard linear system, we may suppose that $\mathcal{P}$ is a standard polynomial system of weight $w(\mathcal{P}) = w$ that is non-linear. As in the linear case, we begin by inserting some additional averaging over a variable m running over an interval of length $\sqrt{N}$. Thus we evaluate the expression

$$\mathcal{I}_\mathcal{P}^* = \int \mathop{\mathbb{E}}_{n \leqslant N} a(n) g_0(x) \prod_{j=1}^k T^{P_j(Wn)} g_j(x) \, d\mu.$$

The argument leading from (3.9) to (3.10) may now be applied, without modification, to show that $\mathcal{I}_{\mathcal{P}}^* = \mathcal{I}_{\mathcal{P}} + o_L(1)$, where

$$\mathcal{I}_{\mathcal{P}}^2 \ll_L \underset{n \leqslant N}{\mathbb{E}} \int \underset{m,m' \leqslant \sqrt{N}}{\mathbb{E}} a(n+m)\overline{a}(n+m')$$

$$\times \prod_{j=1}^k T^{P_j(W(n+m))} g_j(x) \prod_{j=1}^k T^{P_j(W(n+m'))} \overline{g}_j(x)\, d\mu.$$

Next, applying the invariance of μ under the action of T, we find that

$$\mathcal{I}_{\mathcal{P}}^2 \ll_L \underset{n \leqslant N}{\mathbb{E}} \int \underset{m,m' \leqslant \sqrt{N}}{\mathbb{E}} a(n+m)\overline{a}(n+m') g_1(x) \prod_{j=2}^k T^{(P_j - P_1)(W(n+m))} g_j(x)$$

$$\times \prod_{j=1}^k T^{P_j(W(n+m')) - P_1(W(n+m))} \overline{g}_j(x)\, d\mu. \qquad (3.11)$$

When q is an integer, define the set of polynomials $\mathcal{R}_0'(q, q')$ by

$$\mathcal{R}_0'(q, q') = \{P_i(W(n+q')) - P_1(W(n+q))\}_{1 \leqslant i \leqslant k}.$$

Consider the set of polynomials $\mathcal{R}' = \mathcal{R}_0'(m, m) \cup \mathcal{R}_0'(m, m')$, and let

$$\mathcal{R} = \{R_1, \ldots, R_J\}$$

denote the system obtained from $\mathcal{R}'$ by removing the polynomials in $\mathcal{R}'$ of degree zero with respect to n. Then by Lemma B.1, the set of polynomials $\mathcal{R}$ has lower weight with respect to n than the set $\mathcal{P}$, so that $w(\mathcal{R}) < w$. Moreover, the estimate (3.11) takes the shape

$$\mathcal{I}_{\mathcal{P}}^2 \ll_L \underset{n \leqslant N}{\mathbb{E}} \int h_0(x) \underset{m,m' \leqslant \sqrt{N}}{\mathbb{E}} a(n+m)\overline{a}(n+m') \prod_{j=1}^J T^{R_j(m,m',n)} h_j(x)\, d\mu,$$

in which h_j satisfies $\|h_j\|_\infty = O_L(1)$ $(0 \leqslant j \leqslant J)$. Therefore, by the inductive hypothesis in combination with Lemma 3.3, we may conclude that

$$\mathcal{I}_{\mathcal{P}}^2 \ll_L \underset{m,m' \leqslant \sqrt{N}}{\mathbb{E}} (\|a(n+m)\overline{a}(n+m')\|_{V_{\mathcal{R}}} + o(1)) \ll \|a(n)\|_{V_{\mathcal{P}}}^2 + o(1).$$

This confirms the inductive hypothesis for all standard polynomial systems of weight w, and hence the inductive hypothesis (3.8) has now been established for all polynomial systems. $\qquad \square$

Corollary 3.12. *Provided that the lower bound* (3.5) *holds in the special case wherein* X *is an inverse limit of nilsystems of bounded step, then it holds also without restriction.*

Proof. Let Y be the factor supplied by Proposition 3.11, and let $\pi : X \to Y$ be the associated projection. Decompose the characteristic function 1_A by means

of the identity $1_A = \pi^* \pi_* 1_A + (1_A - \pi^* \pi_* 1_A)$. Then one sees that

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) 1_A(x) \prod_{j=1}^{k} (T^{P_j(Wn)} 1_A(x)) \, d\mu \right|$$

$$= \left| \int \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \pi_* 1_A(x) \prod_{j=1}^{k} (T^{P_j(Wn)} \pi_* 1_A(x)) \, d\pi_* \mu \right|$$

$$+ o_w(1) + o_{w \to \infty}(1).$$

$\square$

We make one further reduction, from an inverse limit of nilsystems to a nilsystem proper, and replace 1_A with a Lipschitz continuous function. This we accomplish by means of a standard approximation argument, obtaining a conclusion independent of w.

Lemma 3.13. *Suppose that* X *is an ergodic measure preserving system. Let* $f_1, \ldots, f_k \in L^\infty(X)$, *and suppose that* $\|f_j\|_\infty \leqslant 1$. *If* $g \in L^\infty(X)$ *satisfies the condition* $\|f_1 - g\|_2 < \varepsilon$, *then*

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{b,w}(n) T^{P_1(Wn)}(f_1 - g)(x) \prod_{j=2}^{k} T^{P_j(Wn)} f_j(x) \, d\mu \right| = O_w(\varepsilon).$$

Proof. By applying the triangle inequality in combination with the Cauchy-Schwarz inequality, we obtain

$$\left| \int \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{b,w}(n) T^{P_1(Wn)}(f_1 - g)(x) \prod_{j=2}^{k} T^{P_j(Wn)} f_j(x) \, d\mu \right|$$

$$\leqslant \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{b,w}(n) \left| \int T^{P_1(Wn)}(f_1 - g)(x) \prod_{j=2}^{k} T^{P_j(Wn)} f_j(x) \, d\mu \right|$$

$$\leqslant \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{b,w}(n) \|f_1 - g\|_2 \prod_{j=2}^{k} \|f_j\|_\infty < \varepsilon(1 + o_w(1)).$$

$\square$

We are consequently able to conclude as follows.

Corollary 3.14. *Provided that the lower bound*

$$\mathop{\mathbb{E}}_{n \leqslant N} \int \widetilde{\Lambda}_{w,b}(n) g(x) \prod_{j=1}^{k} (T^{P_j(Wn)} g(x)) \, d\mu \gg_\delta 1 + o_w(1) + o_{w \to \infty}(1)$$

holds in the special case wherein X *is a nilsystem, and* g *is a Lipschitz continuous function satisfying the hypotheses of Theorem 3.9, then it holds also without restriction.*

Now let $X = (G/\Gamma, \mathcal{B}, \mu, T)$ be an ergodic nilsystem, with the transformation T being given by $a \in G$, and write G^0 for the identity component of G. If X is disconnected, then $X^0 = G^0\Gamma/\Gamma \cong G^0/(\Gamma \cap G^0)$ is a connected component of X. Since X is compact, one finds that X is a disjoint union of finitely many translations of X^0, say $X = \cup_{i=1}^J a^i X^0$, and the nilsystem $X_0 = (X^0, a^J)$ has no finite factors. We may now assume further, without loss of generality, that the system X_0 is of the form $(L/\Lambda, b)$, where L is connected and simply connected (see the discussion at the beginning of section 1 of Leibman [13]). In L there is an element c with $c^J = b$. Since a^i induces an isomorphism between X_0 and $X_i = (a^i X^0, a^J)$, the same holds for X_i.

For any Lipschitz continuous function f, and any fixed $x \in X$, the sequence $g_{x,w}(n)$, defined by

$$g_{x,w}(n) = f(a^{P_1(Wn)}x) \ldots f(a^{P_k(Wn)}x), \tag{3.12}$$

is a polynomial nilsequence. Note that for a fixed integer i with $1 \leqslant i \leqslant J$, the set

$$\{(a^{P_1(Wn)}x, \ldots, a^{P_k(Wn)}x)\}_{n \equiv i \ (\mathrm{mod}\ J)}$$

is contained in a fixed connected component of $(G/\Gamma)^k$. Furthermore, if $n = mJ + i$, then

$$a^{P_l(Wn)}x = a^{JP_l'(Wm)+q}x = b^{P_l'(Wm)}a^q x = c^{JP_l'(Wm)+q}c^{-q}a^q x = c^{P_l(Wn)}c^{-q}a^q x,$$

where P_l' and q may depend on i, W, l, J. Thus $f(a^{P_l(Wn)}x)$ can be viewed as a polynomial nilsequence on the nilmanifold L/Λ.

Now consider

$$g_{x,w,i}(n) = 1_{n \equiv i \ (\mathrm{mod}\ J)} f(a^{P_1(Wn)}x) \ldots f(a^{P_k(Wn)}x).$$

The function $1_{n \equiv i \ (\mathrm{mod}\ J)}$ is a 1-step nilsequence on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. It is defined by the polynomial $g : \mathbb{Z} \to \mathbb{R}$ given by $g(n) = n/J$, and a function $F : \mathbb{T} \to [0,1]$ which is Lipschitz, supported on a $1/(10J)$ neighborhood of i/J, and for which $F(i/J) = 1$. Thus $g_{x,w,i}$ is a polynomial nilsequence on the product $\mathbb{T}$ and a connected component of $(G/\Gamma)^k$, with new Lipschitz constant that may depend also on J.

The upshot of the above discussion is that

$$g_{x,w}(n) = \sum_{i=1}^J g_{x,w,i}(n),$$

and thus $g_{x,w}(n)$ can be viewed as a polynomial nilsequence on a nilmanifold G/Γ, where the group G is connected and simply connected.

Proposition 3.15. *With the notation and assumptions in the preamble, one has*

$$\left| \mathop{\mathbb{E}}_{n \leqslant N} (\widetilde{\Lambda}_{w,b}(n) - \eta \widetilde{1}(n)) g_{x,w}(n) \right| = o_{w, \|f\|_{\mathrm{Lip}}}(1). \tag{3.13}$$

This proposition is essentially the polynomial version of Proposition 11.3 of [8]. We sketch a proof below. One of the main ingredients is the following lemma, which is the polynomial version of Proposition 11.2 of [8]. Since the

proof is essentially the same, we omit it, though we note that one could also prove this lemma using Proposition 11.2 of [8] and the fact that a polynomial nilsequence can be viewed as a linear nilsequence on some nilmanifold of larger nilpotence degree. This is shown in Leibman [12] in the context of continuous nilsequences. All that would be required is to verify that Leibman's proof is valid for Lipschitz nilsequences, and is independent of W.

Lemma 3.16. *Let $F(n)$ be a polynomial nilsequence in G/Γ defined by polynomials from the set $\mathcal{P} = \{P_1, \ldots, P_k\}$, as in (3.12), and suppose that F has Lipschitz constant M. Let ε be a positive number, and suppose that $N \geqslant 1$. Then there exist integers $r(\mathcal{P})$ and $t(\mathcal{P})$, and a decomposition $F(n) = F_1(n) + F_2(n)$, where F_1 is an averaged nilsequence on $(G/\Gamma)^{t(\mathcal{P})}$ with Lipschitz constant $O_{M,\varepsilon,G/\Gamma}(1)$, and satisfying*

$$\|F_1(n)\|_{U^{r(\mathcal{P})}[N]_*} \ll_{M,\varepsilon,G/\Gamma} 1,$$

and where $\|F_2\|_\infty = O(\varepsilon)$.

We now replace $\widetilde{\Lambda}_{w,b}(n)$ by

$$\Lambda_{w,b}(n) = \eta \frac{\phi(W)}{W} \Lambda(Wn + b)\widetilde{1}(n),$$

where Λ is the classical von Mangoldt function. This is permissible for averaging purposes in view of the fact that the difference is negligible on average. To this end, we follow section 12 of [8]. Define the function $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ by putting $\chi(x) = x$. We decompose χ via the identity $\chi = \chi^\sharp + \chi^\flat$, where $\chi^\sharp$ is a smooth function vanishing for $|x| \geqslant 1$, and $\chi^\flat$ a smooth function vanishing for $|x| \leqslant \frac{1}{2}$. This induces a decomposition $\Lambda = \Lambda^\sharp + \Lambda^\flat$, with

$$\Lambda^\sharp(n) = -\log R \sum_{d|n} \mu(d)\chi^\sharp\left(\frac{\log d}{\log R}\right)$$

and

$$\Lambda^\flat(n) = -\log R \sum_{d|n} \mu(d)\chi^\flat\left(\frac{\log d}{\log R}\right).$$

Recall the definition of $g_{x,w}(n)$, and define F_x by means of the relation $F_x(Wn) = g_{x,w}(n)$. Let $\varepsilon > 0$ be sufficiently small, and apply Lemma 3.16 to decompose $F_x(Wn)$ in the form $F_{x,1}(Wn) + F_{x,2}(Wn)$, with conditions silently implied by the suffices 1 and 2. From here, following the argument of [8] in order to accommodate the harmless additional factor $\widetilde{1}(n)$, one finds that

$$\left| \mathbb{E}_{n \leqslant N} \left(\frac{\phi(W)}{W} \Lambda^\sharp(Wn + b) - 1 \right) \widetilde{1}(n) F_{x,1}(Wn) \right|$$

$$\leqslant \left\| \left(\frac{\phi(W)}{W} \Lambda^\sharp(Wn + b) - 1 \right) \widetilde{1}(n) \right\|_{U^{r(\mathcal{P})}[N]} \|F_{x,1}(Wn)\|_{U^{r(\mathcal{P})}[N]_*}$$

$$= o_{w,\varepsilon,\|f\|_{\mathrm{Lip}}}(1).$$

On the other hand, in view of the upper bound $\|F_{x,2}(Wn)\|_\infty < \varepsilon$, we see that

$$\left| \underset{n\leqslant N}{\mathbb{E}} \left(\frac{\phi(W)}{W} \Lambda^\sharp(Wn+b) - 1 \right) \widetilde{1}(n) F_{x,2}(Wn) \right|$$

$$\ll \varepsilon \underset{n\leqslant N}{\mathbb{E}} \left| \left(\frac{\phi(W)}{W} \Lambda^\sharp(Wn+b) - 1 \right) \widetilde{1}(n) \right| \ll_w \varepsilon.$$

Taking ε now to be a positive function of N decreasing to zero sufficiently slowly, it follows from the triangle inequality that

$$\left| \underset{n\leqslant N}{\mathbb{E}} \left(\frac{\phi(W)}{W} \Lambda^\sharp(Wn+b) - 1 \right) \widetilde{1}(n) F_x(Wn) \right| = o_{w,\|f\|_{\mathrm{Lip}}}(1). \tag{3.14}$$

For the remaining part of the dissection, we apply Theorem 1.1 of [9]. The sequence

$$h_{x,w}(n) = g_{x,w}((n-b)/W)$$

is a polynomial nilsequence on the same group with the same Lip constant (with a polynomial sequence depending on W), and in addition is of the same degree. Moreover, one has

$$\underset{n\leqslant N}{\mathbb{E}} \Lambda^\flat(Wn+b) g_{x,w}(n) \widetilde{1}(n) = \underset{\substack{b<n\leqslant NW+b \\ n\equiv b \,(\mathrm{mod}\ W)}}{\mathbb{E}} \Lambda^\flat(n) h_{x,w}(n) \widetilde{1}((n-b)/W).$$

The average on the left hand side of (3.13) may therefore be successfully estimated by showing that

$$\log R \underset{m\leqslant \frac{1}{2}NW+b}{\mathbb{E}} \underset{\substack{d\leqslant(\frac{1}{2}NW+b)/m \\ md\equiv b \,(\mathrm{mod}\ W)}}{\mathbb{E}} \mu(d)\chi^\flat\left(\frac{\log d}{\log R}\right) h_{x,w}(md) = o_{w,\|f\|_{\mathrm{Lip}}}(1). \tag{3.15}$$

Fortunately, Theorem 1.1 of [9] implies a bound of the shape

$$\underset{\substack{d\leqslant[M] \\ md\equiv b \,(\mathrm{mod}\ W)}}{\mathbb{E}} \mu(d) h_{x,w}(md) \ll (1+\|f\|_{\mathrm{Lip}})(\log(M/W))^{-A},$$

valid for any positive number A and $M \geqslant 2W$. We note, in particular, that this bound is independent of the polynomial sequence (it depends only on the degree), and that there is no restriction on the size of the coefficients of the latter polynomial. Since the weight $\chi^\flat(x)$ vanishes for $|x| \leqslant \frac{1}{2}$, it follows that the average over d in (3.15) makes no contribution when $[(\frac{1}{2}NW+b)/m] < R^{1/2}$. Consequently, since the weight $\chi^\flat(x)$ is smooth, we find by partial summation that when $m \in [\frac{1}{2}NW+b]$, the inner average on the left hand side of (3.15) is equal to

$$\underset{\substack{R^{1/2}<d\leqslant(\frac{1}{2}NW+b)/m \\ md\equiv b \,(\mathrm{mod}\ W)}}{\mathbb{E}} \mu(d)\chi^\flat\left(\frac{\log d}{\log R}\right) h_{x,w}(md) \ll_w (1+\|f\|_{\mathrm{Lip}})(\log R)^{-A}.$$

We therefore deduce that

$$\underset{m\leqslant \frac{1}{2}NW+b}{\mathbb{E}} \underset{\substack{d\leqslant(\frac{1}{2}NW+b)/m \\ md\equiv b \,(\mathrm{mod}\ W)}}{\mathbb{E}} \mu(d)\chi^\flat\left(\frac{\log d}{\log R}\right) h_{x,w}(md) \ll_w (1+\|f\|_{\mathrm{Lip}})(\log R)^{-A},$$

and, provided that we take $A > 1$, this suffices to deliver the estimate claimed in (3.15).

The conclusion of Proposition 3.15 is obtained by combining the conclusions of (3.14) and (3.15). From here, in view of Corollary 3.14, the lower bound (3.5) follows on noting that by Theorem 3.9, one has

$$\mathop{\mathbb{E}}_{n \leqslant N} \widetilde{1}(n) \int 1_A(x) \prod_{j=1}^{k} (T^{P_j(Wn)} 1_A(x)) \, d\mu \gg_\delta 1 + o(1).$$

The lower bound (3.4) now follows from Lemma 3.8, and this completes the proof of Theorem 1.1.

4. Convergence of multiple averages along the primes

In this section we prove the L^2 convergence of polynomial multiple averages along the primes. Let $f_1, \ldots, f_k$ be bounded functions. Consider the averages

$$A_N(x) = \mathop{\mathbb{E}}_{\substack{p < N \\ p \text{ prime}}} \prod_{j=1}^{k} T^{P_j(p)} f_j(x).$$

We seek to show that the sequence $\{A_N(x)\}_{N=1}^{\infty}$ forms a Cauchy sequence in L^2. Observe first that the sequence $\{A_N(x)\}_{N=1}^{\infty}$ is Cauchy if and only if the sequence $\{B_N(x)\}_{N=1}^{\infty}$ is Cauchy, where

$$B_N(x) = \mathop{\mathbb{E}}_{n < N} \Lambda(n) \prod_{j=1}^{k} T^{P_j(n)} f_j(x).$$

Indeed, independently of the value of x, one has

$$|B_N(x) - A_N(x)| = \left| \mathop{\mathbb{E}}_{n < N} \Lambda(n) \prod_{j=1}^{k} T^{P_j(n)} f_j(x) - \mathop{\mathbb{E}}_{\substack{p < N \\ p \text{ prime}}} \prod_{j=1}^{k} T^{P_j(p)} f_j(x) \right|$$

$$= o(1).$$

It remains now only to show that the sequence $\{B_N(x)\}_{N=1}^{\infty}$ is Cauchy, and this we achieve by applying a stronger version of Proposition 3.11 that we now briefly pause to establish. This may be regarded as a result on orthogonality to nilsystems in L^2.

Proposition 4.1. *Suppose that* X *is an ergodic measure preserving system. Let* $f_1, \ldots, f_k \in L^\infty(X)$, *and suppose that* $\|f_j\|_\infty \leqslant L$ $(1 \leqslant j \leqslant k)$. *Then there exists a factor* Y *of* X, *with* $\pi : X \to Y$ *the factor map, and an integer* $d(\mathbf{P})$, *with the following properties:*

(i) *the factor* Y *has the structure of an inverse limit of* $d(\mathbf{P})$-*step nilsystems, and*

(ii) *if for some index* i *one has* $\pi_* f_i = 0$, *then, uniformly in* b, *one has*

$$\left\| \mathop{\mathbb{E}}_{n \leqslant N} \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) \right\|_2 = o_{L,w}(1) + o_{L,w \to \infty}(1).$$

Proof. In order to establish the proposition, it suffices to confirm that the expression

$$M = \left\| \underset{n \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) \right\|_2^2$$

satisfies the asymptotic relation

$$M = o_{L,w}(1) + o_{L,w\to\infty}(1).$$

But if we write

$$g_n(x) = \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x),$$

then an application of the triangle inequality to the expansion of M yields

$$M = \underset{n,m \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \widetilde{\Lambda}_{w,b}(m) \int \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) \prod_{j=1}^{k} T^{P_j(Wm+b)} \overline{f}_j(x) \, d\mu$$

$$\leqslant \underset{n \leqslant N}{\mathbb{E}} \, \widetilde{\Lambda}_{w,b}(n) \left| \underset{m \leqslant N}{\mathbb{E}} \int \widetilde{\Lambda}_{w,b}(m) g_n(x) \prod_{j=1}^{k} T^{P_j(Wm+b)} \overline{f}_j(x) \, d\mu \right|. \qquad (4.1)$$

However, the hypotheses of the proposition imply that $\|g_n(x)\|_\infty \leqslant L^k$, and so it follows from Proposition 3.11 that, uniformly in b, one has

$$\left| \underset{m \leqslant N}{\mathbb{E}} \int \widetilde{\Lambda}_{w,b}(m) g_n(x) \prod_{j=1}^{k} T^{P_j(Wm+b)} \overline{f}_j(x) \, d\mu \right| = o_{L,w}(1) + o_{L,w\to\infty}(1).$$

The desired conclusion now follows on substituting this estimate into (4.1). $\qquad \square$

We now return to the proof of Theorem 1.2. Suppose that $f_1, \ldots, f_k \in L^\infty(X)$, and that $\|f_j\| \leqslant L$ $(1 \leqslant j \leqslant k)$. Let M be a large natural number, and put $N = 2M$. Observe that since the von Mangoldt function Λ is supported on prime powers, one has

$$B_{WM}(x) = \underset{n < WM}{\mathbb{E}} \, \Lambda(n) \prod_{j=1}^{k} T^{P_j(n)} f_j(x)$$

$$= \frac{1}{\phi(W)} \sum_{\substack{0 \leqslant b < W \\ (b,W)=1}} \underset{n < M}{\mathbb{E}} \, \frac{\phi(W)}{W} \Lambda(Wn+b) \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) + o_L(1)$$

$$= \frac{1}{\phi(W)} \sum_{\substack{0 \leqslant b < W \\ (b,W)=1}} \underset{n < N}{\mathbb{E}} \, \eta^{-1} \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) + o_{L,w}(1).$$

By applying the triangle inequality in combination with Proposition 4.1, we obtain

$$\left\| \mathop{\mathbb{E}}_{n<N} \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x) - \mathop{\mathbb{E}}_{n<N} \widetilde{\Lambda}_{w,b}(n) \prod_{j=1}^{k} T^{P_j(Wn+b)} \pi^* \pi_* f_j(x) \right\|_2$$
$$= o_{L,w}(1) + o_{L,w\to\infty}(1),$$

in which π is the projection onto the relevant nilpotent factor supplied by Proposition 3.11. By Proposition 3.15, meanwhile, one has

$$\left\| \mathop{\mathbb{E}}_{n<N} (\eta^{-1}\widetilde{\Lambda}_{w,b}(n) - \widetilde{1}(n)) \prod_{j=1}^{k} T^{P_j(Wn+b)} \pi^* \pi_* f_j(x) \right\|_2 = o_{L,w}(1).$$

Consider next the average $C_{b,M}(x)$ defined by

$$C_{b,M}(x) = \mathop{\mathbb{E}}_{n<M} \prod_{j=1}^{k} T^{P_j(Wn+b)} f_j(x).$$

It follows from Leibman [11] that the sequence $\{C_{b,M}(x)\}_{M=1}^{\infty}$ converges, and is thus a Cauchy sequence. Fix a positive number ε. Then whenever M_1 and M_2 are sufficiently large, one has

$$\|C_{b,M_1}(x) - C_{b,M_2}(x)\|_2 \leqslant \varepsilon.$$

Under the same conditions, moreover, it follows from the triangle inequality in combination with the conclusions of the previous paragraph that

$$\left\| B_{WM_i}(x) - \frac{1}{\phi(W)} \sum_{\substack{0 \leqslant b < W \\ (b,W)=1}} C_{b,M_i}(x) \right\|_2 < \varepsilon \quad (i = 1, 2).$$

Consequently, again by the triangle inequality, one finds that whenever M_1 and M_2 are sufficiently large, one has $\|B_{WM_1}(x) - B_{WM_2}(x)\|_2 \leqslant 3\varepsilon$, and thus the sequence $\{B_{WM}(x)\}_{M=1}^{\infty}$ is Cauchy. Finally, since for $1 \leqslant i \leqslant W$, one has

$$B_{WM+i}(x) = B_{WM}(x) + o_L(1),$$

the sequence $\{B_M(x)\}_{M=1}^{\infty}$ is also Cauchy. This confirms our earlier claim, and thus the proof of Theorem 1.2 is complete.

Appendix A. Ergodic theoretic preliminaries

We take the opportunity here to prepare some of the infrastructure central to the ergodic theory employed in the main body of this paper.

A.1. **Measure preserving systems.** We begin by recalling that a *measure preserving transformation* on a measure space $(X_0, \mathcal{B}_X, \mu_X)$ is a map $T : X_0 \to X_0$ satisfying the property that, for all $B \in \mathcal{B}$, one has $\mu(T^{-1}B) = \mu(B)$. A *probability measure preserving system (m.p.s.)* X is a quadruplet $(X_0, \mathcal{B}_X, \mu_X, T)$, where the triple $(X_0, \mathcal{B}_X, \mu_X)$ is a probability measure space, and $T : X_0 \to X_0$ a measure preserving transformation. We define the L^p spaces $L^p(X) = L^p(X_0, \mathcal{B}_X, \mu_X)$ for $1 \leqslant p \leqslant \infty$ in the usual manner. Thus, in

particular, we identify any two functions in $L^p(X)$ which agree μ-almost everywhere. If X_0 is a point, we write $X = \mathrm{pt}$. We will assume throughout this paper that the measure preserving system X is *regular*, which is to say that X_0 is a compact metric space and $\mathcal{B}_X$ consists of all Borel sets in X. There is no loss of generality in this assumption since any m.p.s X such that $\mathcal{B}_X$ is generated by a countable set is equivalent to a regular one.

A *factor map* $\pi_Y^X : X \to Y$ is a morphism in the category of measure preserving systems. A *factor* $(Y_0, \mathcal{B}_Y, \mu_Y, S, \pi_Y^X)$ of a system $X = (X_0, \mathcal{B}_X, \mu_X, T)$ is a measure preserving system $Y = (Y_0, \mathcal{B}_Y, \mu_Y, S)$ together with a factor map $\pi_Y^X : X \to Y$. In these circumstances, the pushforward $(\pi_Y^X)_* \mu_X$ is equal to μ_Y, and the relation $\pi_Y^X \circ T = S \circ \pi_Y^X$ holds μ_X-almost everywhere. When $f : Y \to \mathbb{C}$ is a measurable map, we write $(\pi_Y^X)^* f : X \to \mathbb{C}$ for the pullback defined by $(\pi_Y^X)^* f = f \circ \pi_Y^X$. Conversely, when $f \in L^2(X)$, we denote by $(\pi_Y^X)_* f \in L^2(Y)$ the pushforward of f. We then define the conditional expectation of f to Y by

$$\mathbb{E}(f|Y) = (\pi_Y^X)^*(\pi_Y^X)_* f \in L^2(X).$$

We say that $f \in L^2(X)$ is *Y-measurable* when $f = \mathbb{E}(f|Y)$, or equivalently, when $f = (\pi_Y^X)F$ for some $F \in L^2(Y)$. In circumstances wherein Y_0 is a point, we say that Y is *trivial* and denote Y as pt. Thus, for instance, we may write $(\pi_{\mathrm{pt}}^X)_* f = \int_X f \, d\mu_X$. When there is no ambiguity we write π for π_Y^X. It is convenient when confusion is readily avoided to abuse notation by writing X for the system $(X_0, \mathcal{B}_X, \mu_X, T)$, or for the measure space $(X_0, \mathcal{B}_X, \mu_X)$, or simply for the phase space X_0.

A.2. **Nilsystems and nilsequences.** A *k-step nilsystem* X is a measure preserving system $(X_0, \mathcal{B}_X, \mu_X, T)$, in which $X_0 = G/\Gamma$, for some k-step nilpotent Lie group G and a cocompact lattice Γ, and $\mathcal{B}_X$ is the Borel σ-algebra, μ_X the Haar measure, and the measure preserving transformation $T : G/\Gamma \to G/\Gamma$ is given by a rotation by some group element $a \in G$, which is to say that $T(g\Gamma) = ag\Gamma$. A *k-step (linear) nilsequence* is a sequence of the form $\{F(a^n x)\}_{n \in \mathbb{N}}$, where $x \in G/\Gamma$ and $F : G/\Gamma \to \mathbb{R}$ is a continuous function. We endow the nilmanifold G/Γ with a smooth Riemannian metric d. Let $g : \mathbb{N} \to G$. For $h \in \mathbb{N}$ we denote $\partial_h g(n) = g(n+h)g^{-1}(n)$. A function $g : \mathbb{N} \to G$ is called a polynomial sequence of degree $< k$ if, for any $h_1, \ldots, h_k \in \mathbb{N}$, one has $\partial_{h_k} \ldots \partial_{h_1} g(n) \equiv 1_G$. A *degree $< k$ polynomial nilsequence* is a sequence of the form $\{F(g(n)x)\}_{n \in \mathbb{N}}$, where $x \in G/\Gamma$ and $F : G/\Gamma \to \mathbb{R}$ is a continuous function, and g is a polynomial sequence of degree $< k$. We say that a nilsequence $\{F(g(n)x)\}$ has Lipschitz constant L if the function F has Lipschitz constant L. In circumstances in which the representation of the nilsequence is not explicit, we define the Lipschitz constant by taking the infimum over all possible representations.

We next define the Gowers norms, introduced in Lemma 3.9 of [7]. Let a be a function from $\mathbb{Z}/N\mathbb{Z}$ into $\mathbb{C}$. When k is a non-negative integer, we define

the $U^k[N]$-*norm* of a to be the quantity $\|a\|_{U^k[N]}$ defined via the relation

$$\|a\|_{U^k[N]}^{2^k} = \mathop{\mathbb{E}}_{n, m_1, \ldots, m_k \in \mathbb{Z}/N\mathbb{Z}} \prod_{\boldsymbol{\omega} \in \{0,1\}^k} a^{\boldsymbol{\omega}}(n + \boldsymbol{\omega} \cdot \mathbf{m}),$$

where $a^{\boldsymbol{\omega}} = a$ when $\sum_{i=1}^{k} \omega_i \equiv 0 \pmod{2}$, and otherwise $a^{\boldsymbol{\omega}} = \overline{a}$. Next, we define the *dual norm to the Gowers norm* by means of the relation

$$\|F\|_{U^k[N]^*} = \sup\left\{ \left| \mathop{\mathbb{E}}_{n \in [N]} f(n)F(n) \right| : \|f\|_{U^k[N]} \leqslant 1 \right\}.$$

Finally, an *averaged k-step nilsequence* with Lipschitz constant M is a function $F(n)$ having the form

$$F(n) = \mathop{\mathbb{E}}_{i \in I} F_i(a_i^n x_i),$$

where I is a finite index set, and for each $i \in I$, the expression $F_i(a_i^n x_i)$ is a bounded k-step nilsequence on G/Γ with Lipschitz constant not exceeding M.

APPENDIX B. PET INDUCTION

The notion of PET induction was introduced by Bergelson in [1] as a mechanism for establishing a *Polynomial Ergodic Theorem* (or PET) for a weakly mixing system. We introduce the framework required to apply PET induction in this appendix so as to assist in our exposition elsewhere in this paper.

A *polynomial system* is a set of polynomials $\mathcal{P} = \{P_1(n), \ldots, P_k(n)\}$, where $P_i(n) \in \mathbb{Z}[n]$ $(1 \leqslant i \leqslant k)$. The *degree* of $\mathcal{P}$ is the maximum of the degrees of the polynomials lying in $\mathcal{P}$. We define an equivalence relation on $\mathbb{Z}[n]$ by defining the polynomials P and Q to be equivalent when $\deg(P - Q) < \deg P$. We then define the *degree* of an equivalence class to be the degree of its elements. Any polynomial system $\mathcal{P}$ can be partitioned into equivalence classes. For each positive integer l, let w_l be the number of classes of degree l in $\mathcal{P}$. Then the *weight $w(\mathcal{P})$* of the system $\mathcal{P}$ is defined to be the vector $(w_1, \ldots, w_{\deg \mathcal{P}})$. Next we establish an order relation on weight vectors. Given two integer vectors $\mathbf{v} = (v_1, \ldots, v_r)$ and $\mathbf{w} = (w_1, \ldots, w_s)$, we write $\mathbf{v} < \mathbf{w}$ if either $r < s$, or else $r = s$ and there is an index n for which $v_j = w_j$ $(n < j \leqslant r)$ and $v_n < w_n$. Subject to this relation, the set of weights of polynomial systems is well-ordered. The PET induction is an induction on this well-ordered set.

An ordered system $\mathcal{P} = \{P_1 \ldots, P_k\}$ is *standard* if $\deg P_j > 0$ for $1 \leqslant i \leqslant k$, $\deg(P_i - P_j) > 0$ for $i \neq j$, and in addition P_1 has minimal degree in $\mathcal{P}$. The system is *linear* if each polynomial in $\mathcal{P}$ is linear. The following lemma shows that standard systems are well-behaved with respect to a natural differencing operation.

Lemma B.1. *Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be an ordered polynomial system satisfying the property that P_1 has minimal degree in $\mathcal{P}$. Given a positive integer m, let $\mathcal{Q}'_m$ be the system defined by*

$$\mathcal{Q}'_m = \{P_j(n + m) - P_1(n) : 1 \leqslant j \leqslant k\},$$

and let $\mathcal{Q}_m^*$ be the set of polynomials lying in $\mathcal{Q}_m' \cup \mathcal{Q}_0'$ having degree zero in terms of n. Finally, denote by $\mathcal{Q} = \mathcal{Q}_m(\mathcal{P})$ the system obtained from the set $(\mathcal{Q}_m' \cup \mathcal{Q}_0') \setminus \mathcal{Q}_m^*$ by reordering, if possible, so as to respect the conditions described in the preamble. Then

  (i) when $\mathcal{P}$ is standard and non-linear, the system $\mathcal{Q}$ is standard of weight strictly smaller than the weight of $\mathcal{P}$, and

  (ii) when $\mathcal{P}$ is linear, the system $\mathcal{Q}$ is of weight strictly smaller than the weight of $\mathcal{P}$, though possibly non-standard.

Note that when $\mathcal{P}$ is a system of weight (1), then the system $\mathcal{Q}$, associated to $\mathcal{P}$ by the lemma, is empty. Given a standard polynomial system $\mathcal{P}$, the number of steps of the type described in the lemma required to reach the empty system is called the *parallelepiped degree* of $\mathcal{P}$, denoted by $l(\mathcal{P})$.

Example B.2. Consider the situation in which $\mathcal{P} = \{n^2, n^2 + n\}$. Then $\mathcal{P}$ is standard of weight $(0, 1)$. The system $\mathcal{Q}_m(\mathcal{P})$ associated to $\mathcal{P}$ by Lemma B.1 is the system $\mathcal{P}_1 = \{n, 2mn + m^2, 2mn + n + m^2 + m\}$, which is standard of weight (3). A second application of the lemma associates the system $\mathcal{P}_2 = \mathcal{Q}_k(\mathcal{P}_1)$ to $\mathcal{P}_1$, and this is the system

$$\{2mn - n + m^2, 2mn + m^2 + m, 2mn - n + 2km + m^2, 2mn + 2mk + m^2 + m + k\},$$

which is non-standard of weight (2). Another application yields the system $\mathcal{P}_3 = \mathcal{Q}_l(\mathcal{P}_2)$, namely

$$\{n + m, n + 2mk + m + k, n + 2ml + m, n + 2mk + 2ml + m + k\},$$

which is of weight (1). Finally, one last application gives the empty set. Thus we may conclude that $l(\mathcal{P}) = 4$.

References

[1] V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), 337–349.

[2] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.

[3] V. Bergelson, A. Leibman and E. Lesigne, *Intersective polynomials and the polynomial Szemerédi theorem*, Adv. Math. **219** (2008), 369–388.

[4] J. Bourgain, *An approach to pointwise ergodic theorems*, Geometric aspects of functional analysis (1986/87), pp. 204–223, Lecture Notes in Math. **1317**, Springer, Berlin, 1988.

[5] N. Frantzikinakis, B. Host and B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, J. Reine Angew. Math. **611** (2007), 131–144.

[6] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981.

[7] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[8] B. Green and T. Tao, *Linear equations in primes*, Annals of Math. (2) **171** (2010), 1753–1850.

[9] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, preprint.

[10] B. Green, T. Tao and T. Ziegler *An inverse theorem for the Gowers $U(4)$ norm*, Glasgow Math. J. (to appear).

[11] A. Leibman, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel J. Math. **146** (2005), 303-315.

[12] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), 201-213.

[13] A. Leibman, *Multiple polynomial correlation sequences and nilsequences*, Ergodic Theory Dynam. Systems (to appear).

[14] H.-Z. Li and H. Pan, *Difference sets and polynomials of prime variables*, Acta Arith. **138** (2009), 25–52.

[15] J. Lucier, *Difference sets and shifted primes*, Acta Math. Hungar. **120** (2008), 79–102.

[16] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory I. Classical theory*, Cambridge Studies in Advanced Mathematics, 97, Cambridge University Press, Cambridge, 2007.

[17] I. Z. Ruzsa and T. Sanders, *Difference sets and the primes*, Acta Arith. **131** (2008), 281–301.

[18] A. Sárközy, *On difference sets of sequences of integers. III*, Acta Math. Acad. Sci. Hungar. **31** (1978), 355–386.

[19] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), 213–305.

[20] M. Wierdl, *Pointwise ergodic theorem along the prime numbers*, Israel J. Math. **64** (1988), 315–336.

[21] M. Wierdl, *Almost everywhere convergence and recurrence along subsequences in ergodic theory*, Ph.D. Thesis, Ohio State University, 1989.

TDW: School of Mathematics, University of Bristol, University Walk, Clifton, Bristol BS8 1TW, United Kingdom

E-mail address: matdw@bristol.ac.uk

TDZ: Department of Mathematics, Technion-Israel Institute of Technology, Haifa 32000, Israel

E-mail address: tamarzr@tx.technion.ac.il